Psychological Monographs

General and Applied

No. 377
1954

Some Statistical Relationships between Selected Response Dimensions and Problem-Solving Proficiency

By

Nicholas A. Fattu, Edmund Mech, and Ervin Kapos

Institute of Educational Research, Indiana University

Price, $1.00

Fattu, Mech, and Kapos

Vol. 68
No. 6

Edited by Herbert S. Conrad

Published by The American Psychological Association, Inc.

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

Editor

HERBERT S. CONRAD

Department of Health, Education, and Welfare
Office of Education
Washington 25, D.C.

Managing Editor

LORRAINE BOUTHILET

Consulting Editors

DONALD E. BAIER
FRANK A. BEACH
ROBERT G. BERNREUTER
WILLIAM A. BROWNELL
HAROLD E. BURTT
JERRY W. CARTER, JR.
CLYDE H. COOMBS
JOHN G. DARLEY
JOHN F. DASHIELL
EUGENIA HANFMANN
EDNA HEIDBREDER
HAROLD E. JONES
DONALD W. MACKINNON
LORRIN A. RIGGS
CARL R. ROGERS
SAUL ROSENZWEIG
ROSS STAGNER
PERCIVAL M. SYMONDS
JOSEPH TIFFIN
LEDYARD R. TUCKER
JOSEPH ZUBIN

MANUSCRIPTS should be sent to the Editor.

Because of lack of space, the *Psychological Monographs* can print only the original or advanced contribution of the author. *Background and bibliographic materials must, in general, be totally excluded* or kept to an irreducible minimum. Statistical tables should be used to present only the most important of the statistical data or evidence.

The first page of the manuscript should contain the title of the paper, the author's name, and his institutional connection (or his city of residence). Acknowledgments should be kept brief, and appear as a *footnote* on the first page. No table of contents need be included. For other directions or suggestions on the preparation of manuscripts, see: CONRAD, H. S. Preparation of manuscripts for publication as monographs. *J. Psychol.*, 1948, 26, 447-459.

CORRESPONDENCE CONCERNING BUSINESS MATTERS (such as author's fees, subscriptions and sales, change of address, etc.) should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C. Address changes must arrive by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee third-class forwarding postage.

COPYRIGHT, 1954, BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Vol. 68, No. 6 Whole No. 377, 1954

Psychological Monographs: General and Applied

Some Statistical Relationships between Selected Response Dimensions and Problem-Solving Proficiency

Nicholas A. Fattu, Edmund Mech, and Ervin Kapos
Institute of Educational Research, Indiana University

IT IS WELL known that the "higher" thought processes, more generally subsumed under the rubric of problem solving, have not received the experimental attention from psychologists that they merit.

In a recent appraisal of research on thinking, Underwood concludes ". . . that a great deal of the future development in thinking rests squarely on filling the hiatus representing lack of empirical relationships between stimulus variables and response changes for tasks said to be measuring thinking" (5, p. 211).

Contradiction of the foregoing statement by Underwood is indeed difficult. If Underwood's "formula" is correct, clear-cut relationships between environmental variables and specific response changes in problem-solving tasks are a basic requirement.

It might be well at this point to consider the nature of response changes in relation to problem-solving research. Most problem-solving studies compare various groups of problem solvers on the basis of (*a*) number of problems solved correctly and (*b*) time taken to solve a problem. That time is in some cases a dubious criterion is pointed out in recent problem-solving investigations (1, 4).

The use of number of problems correctly solved is, of course, a valid response measure. It is, however, an *end result* which yields no information with reference to describing the behaviors or response categories that are important during the actual problem-solving process.

The lack of sophisticated response measures in problem-solving research is clear. Before environmental manipulations can be related to response measures in problem-solving studies, more adequate response measures must be developed. It is not sufficient to describe the "good" problem solver as simply one who solves the most problems in a given situation.

A primary need in the area of problem solving, then, is an objective description of the *behaviors* exhibited by the more proficient problem solvers, as contrasted with the less proficient problem solvers, while in the actual process of solving problems.

A focal problem in the present paper was to investigate selected response dimensions and determine their relative usefulness in differentiating between problem solvers of varying ability levels. In addition to time measures, it was proposed to study under specified conditions such response concepts as (*a*) patterns of search, (*b*) stereotypy of response, and (*c*) error distance.

It has often been suggested in various

portions of the problem-solving literature that good problem solvers are less stereotyped in their approach than the poorer problem solvers. Proposals have been offered to the effect that pattern of search or "how" the problem solver goes about his task is somehow related to efficiency. This supposition leads directly to the question of whether good problem solvers employ a different search pattern from the less proficient problem solvers.

The error-distance concept appeared worthy of investigation for the following reason. When the term "success" is used with reference to solving a problem, it is implied that the problem solver was absolutely correct. However, it is proposed that degrees of correctness be considered as a possible parameter in problem-solving situations. For example, two persons are required to solve a problem. Both respondents give erroneous solutions when an absolute standard of correctness is applied. However, Problem Solver A is extremely close to a correct solution, while Problem Solver B misses the solution by a wide margin. For this reason, it was decided to make a thorough analysis of whether the more proficient problem solver, although incorrect in his solution, was consistently less in error distance than the less proficient problem solvers.

The foregoing response dimensions of time, pattern of search, response stereotypy, and error distance were studied in two broad groups of problem solvers, classified with respect to proficiency, as being either "good" or "poor" problem solvers. The basic problem was to study the relationship between these particular response indices and problem-solving proficiency.

A subsidiary consideration was to determine to what extent problem solvers of varying proficiency levels modify their "responses" when training and practice factors are common to all problem solvers. Stated another way, we wished to determine the extent to which the response dimensions under study were modifiable under conditions of similar training and practice. It seemed important to learn whether those subjects initially classified as poor problem solvers become *less* stereotyped when compared with their original stereotypy indices.

Simply then, it is the purpose of this paper to describe the behaviors good problem solvers exhibit but that poorer problem solvers apparently lack. Conversely, the poor problem solver may exhibit certain characteristics that the good problem solver lacks. A secondary purpose is to gain an indication of whether the initially poor problem solver tends to resemble the good problem solver when training and practice factors are introduced.

Method

Apparatus

The apparatus used is shown in Figure 1. It is referred to as the gear-train apparatus, designed for exploratory research in the development of problem-solving response measures. The gear train consists of sets of gears and shafts mounted on a sheet of aluminum 1/2 inch thick, 29 inches long, and 20 inches wide. The gear trains were designed so that two series and four parallel channels were formed, providing for crossed information chains.

Two operating control handles, labeled A and B, supplied the input necessary to set the gear train in motion. The motion was transferred through the gear trains and, as an end result, activated a switch that caused a series of red lights to illuminate the control panel.

FIG. 1. The gear-train apparatus.

These red lights would illuminate only if the gear train was working properly and if control A was turned 13 revolutions and control B was turned 13 revolutions.

When the required number of revolutions was made and the end result (illumination of control panel) failed to occur, this indicated to *S* that something was wrong in the gear train. The problem for *S* in each case was to locate the defect or malfunction.

Problems or Malfunctions

For the purposes of this investigation, five classes of problems or malfunctions were utilized: (*a*) a gear not meshing due to a gap between gears, (*b*) a slipping gear due to a set screw having been removed, (*c*) a damaged cog, or broken gear, (*d*) "special" problems, such as defects introduced into one of the differentials, and (*e*) problems referred to as "no-defect" problems, i.e., nothing was actually wrong with the equipment.

Classification of Subjects

The subjects used in this study were 49 adults enrolled in the School of Education at Indiana University during the 1953 summer session. Participation in the experiment was required of all *Ss*.

Each *S* spent approximately 2 hours in the experimental situation.

Procedure

The investigation was divided into three "test" phases: the initial or pretest, intertest, and posttest. Each *S* was given six problems or malfunctions to

solve during each of the three problem-solving phases, making a total of 18 problems that each *S* was confronted with.

The 49 *Ss* were divided into two initial groups on the basis of their performance on the six pretest malfunctions. The first group, Group I, included all *Ss* who successfully solved at least three of the initial six problems; the second group included all others. This latter group was then subdivided into two groups on the basis of the difference between their pretest scores and their posttest scores. Those who demonstrated an improvement of at least two malfunctions between the pretest and the posttest were classified into Group II, or poor "improved" problem solvers. All other *Ss* remained in Group III, or poor "unimproved" problem solvers.

The *Ss* were run individually. Immediately upon entering the experimental situation for the initial phase, *S* was given the standard operating procedure (SOP) for the gear-train apparatus. The SOP consisted of turning control handles A and B, 13 and 13 revolutions respectively; if the control panel failed to illuminate, this indicated to *S* that the equipment was malfunctioning. The required solution in each case was the correct location of the malfunction.

After the SOP orientation, *S* was given a series of six malfunctions to locate. Performance on these six problems constituted a measure of initial ability in this specific situation for each *S*. Each problem in the series was given singly; *S* waited in an adjoining room while each succeeding problem was being inserted into the apparatus by *E*. The *Ss* were allowed 4 minutes to solve each problem. There was an interval of 1 minute after each problem. This was necessary to allow *E* to prepare the next problem for *S*.

The *Ss* were not required to give an answer and, in addition, they were free to give up prior to the 4-minute time limit set on each problem. After *S* had performed on the initial six problems that were administered to determine initial levels of problem-solving ability on the gear train, the first phase of training, in the form of a tape-recorded "basic knowledge" lecture approximately 6 minutes long, was individually administered to each *S*. This lecture was played twice for each *S*. Integrated with the lecture was a series of 35-mm. slides projected on a screen. The rationale of this lecture was to convey the basic nomenclature of the gear train. Included were the concepts of transfer of motion, function of gears, bearings, and shafts. A zero time interval existed between the basic knowledge lecture and the intertest, which consisted of *S*'s being administered a second series of six gear-train problems.

A 24-hour interval existed between the intertest and the third series of problems. Upon returning to the laboratory the following day, *S* was first given additional training in the form of a 7½-minute tape-recorded lecture on principles of problem solving specific to the gear train. This problem-solving lecture was based on a simplified version of a previously developed symptom analysis guide to "trouble-shooting." The lecture, which was played twice for each *S*, stressed starting with the greatest magnitude of error, locating the correctly operating component nearest the greatest error, and once these two points were bracketed, locating the malfunction somewhere between. Following this second training session *S* was given a third

series of six gear-train problems. This was the final series of problems for each *S* and constituted the posttest measure.

All moving parts of the gear train were numbered. These numbers ranged from 1 to 66, with the two operating handles labeled A and B. During each of the 18 problems, *S* was required to call out the numbers of all parts he looked at, as he went from part to part in the process of attempting a correct solution.

Responses were considered as correct only if *S* pointed to the exact location of the malfunction. Time measures were taken to the nearest 5 seconds.

Response Measures

The usual response measures of problems solved and time were taken. Three additional classes of responses in the form of error distance, pattern of search, and behavioral stereotypy were recorded for all *S*s. Descriptions of the three additional response measures are given as follows:

1. *Error distance.* For purposes of computation the gear train was classified into 25 functional units (FU), with each FU consisting of a shaft and all the gears on it, or of a differential. All "error distances" were based on the 25 FU's of the gear train and computed on the basis of the number of "jumps" from FU to FU necessary to travel from the location of the response to the location of the malfunction.

The scoring procedure for error distance was as follows. Correct answers were given an error-distance score of zero. Responses that identified the incorrect part as the malfunction but which were given in the correct functional unit were given error distances of 0.5. The no-defect problems where an imaginary malfunction was "located" presented a special case. In those cases the no-defect problem was assigned to functional units 0 and 26. These may be thought of as the two control handles. Error distances from any response were measured to the farther of these two locations. This method appears defensible because the first task of the gear-train problem solver is to apply the SOP in order to verify the existence of a malfunction. Failure in this preliminary task ranks a subject below one who has succeeded in this task but failed in a subsequent one.

2. *Patterns of search.* The total responses of each group on each of the 25 functional units for each of the 18 problems were tabulated. The 18 problems or malfunctions were inserted in different parts of the 25 functional units, their various locations being determined by random sampling. The basic idea here was to compare how good and poor problem solvers pattern their responses. Clearly, the location of each defect was in one of the 25 functional units. The pattern-of-search measure entailed recording the frequency of occurrence of *S*'s going to each of the 25 FU's while in the process of locating each of the 18 malfunctions.

3. *Stereotypy of response.* For purposes of obtaining stereotypy indices the gear-train apparatus was redivided, for computational purposes, into six logically coherent clusters. There existed several open spaces between clusters or sets of gears, which permitted this convenient breakdown to six areas. It was desired to obtain stereotypy indices for the various groups of problem solvers for each of the three phases of the investigation.

Results

1. *Analysis of correct solutions.* First, an analysis was made of the performance of the 49 *S*s with respect to number of correct solutions for each of the three test phases. These data, condensed in terms of means, standard deviations, and *F* ratios, are shown in Table 1.

The mean scores for each of the proficiency groups were based upon a series of six problems during three test sessions. Thus the average problem solver in Group I (good problem solvers) gave 3.5 correct solutions during the pretest, of a possible six correct solutions.

The total test mean scores are indicative of the relative performances of the three problem-solving groups over the entire series of 18 problems. The remaining section of Table 1 summarizes the analysis of variance for differences among the three problem-solving groups. In addition to the analysis for total test differences, *F* ratios were computed for differences during the pre-, inter-, and posttest phases. The obtained *F* of 42.42 for the pretest phase was significant be-

TABLE 1

COMPARISON OF CORRECT SOLUTIONS MADE BY *Ss* CLASSIFIED ACCORDING TO PROBLEM-SOLVING PROFICIENCY

Means and Standard Deviations

	Group I		Group II		Group III	
Test Phase	Mean	*SD*	Mean	*SD*	Mean	*SD*
Pretest	3.5	0.65	1.2	0.66	1.5	0.72
Intertest	3.0	1.58	1.8	1.28	1.7	0.92
Posttest	3.8	0.83	4.3	1.06	1.9	0.97
Total test	10.3	2.01	7.3	2.43	5.1	2.02

Analysis of Variance Results

Test Phase	Source of Variance	*df*	Mean Square	*F*
Pretest	Between groups	2	21.21	42.42**
	Within groups	46	0.50	
Intertest	Between groups	2	7.38	4.64*
	Within groups	46	1.59	
Posttest	Between groups	2	29.53	29.53**
	Within groups	46	1.00	
Total test	Between groups	2	103.77	21.09**
	Within groups	46	4.92	

* $p < .05$.
** $p < .01$.

yond the 1 per cent level of confidence. It should be made clear that it was necessary to demonstrate at the outset that those classified as good problem solvers (Group I) were decidedly more proficient than those classed as poor problem solvers (Groups II and III). The foregoing *F* ratio of 42.42 establishes statistically that these groups were initially different with respect to ability to solve gear-train problems.

Immediately prior to the second series of problems, the intertest, all *Ss* were given a training session in the form of a basic knowledge lecture concerning the elements of the gear-train apparatus. From the reduced *F* ratio for intertest differences it appears safe to conclude that this lecture had some effect upon the less proficient problem solvers (Groups II and III). The obtained *F* ratio of 4.64 was, nonetheless, significant at the 5 per cent level of confidence. The difference again favored the Group I problem solvers. This basic knowledge lecture probably contained little that Group I *Ss* did not already know and consequently was not a factor in their performance on the series of six intertest problems. On the other hand, Groups II and III made fractional mean gains, indicating that the basic knowledge training session was of some benefit.

Prior to the final series of six gear-train problems, designated as the posttest, all *Ss* received a second training session in the form of a tape-recorded "How to solve gear-train problems" lecture.

The obtained *F* of 29.53, significant beyond the 1 per cent level of confidence, indicated that differences existed among the three groups with respect to posttest problems correctly solved. Group II

TABLE 2

MEANS AND STANDARD DEVIATIONS OF TIME IN SECONDS REQUIRED TO DECIDE THE LOCATION OF SOLUTIONS TO PRE-, INTER-, AND POSTTEST PROBLEMS

Test Phase	Group I		Group II		Group III	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
All Problems						
Pretest	74.7	19.80	108.4	57.54	89.4	56.60
Intertest	61.3	26.07	79.2	51.42	52.6	30.94
Posttest	65.7	28.61	63.7	17.19	53.9	32.54
Total test	67.2	20.22	84.0	36.72	65.4	34.56
Problems Solved						
Pretest	53.4	20.12	66.2	50.70	59.2	49.61
Intertest	40.2	15.29	50.0	29.89	36.4	26.98
Posttest	38.4	15.35	52.4	14.42	39.3	18.63
Total test	44.2	10.33	56.2	16.99	42.2	22.64
Problems Failed						
Pretest	100.1	43.15	115.4	60.47	102.6	68.18
Intertest	95.4	71.06	90.2	58.29	61.2	45.07
Posttest	111.1	69.62	84.2	39.33	64.5	43.76
Total test	93.5	45.17	88.6	39.24	77.6	45.41

(initially poor, but improved) showed the greatest gain in mean number of problems solved when compared with Group I or III. However, the *F* of 21.09 obtained from group performances over the series of 18 problems was statistically significant beyond the 1 per cent level in favor of Group I (good problem solvers).

The consistent superiority of Group I *Ss* over the remaining two groups had at this point been established. The single exception consisted of the posttest performance of Group II *Ss*, but the differences between Group I and II *Ss* were not actually significant at this point. However, Groups I and II were significantly better than Group III on the posttest problems.

The point is being made that the foregoing tests of significance clearly show Group I *Ss* to be sampled from a "different" population of problem solvers than either Group II or III. We are, of course, referring to problem-solving ability as it relates to the gear-train apparatus. We now approach the focal problem with which this paper is concerned. The central objective is the investigation of selected relationships between well-defined response categories and problem-solving proficiency. The response categories that were examined are time, error distance, pattern of search, and behavioral indices of response stereotypy.

What we wish to learn is whether the response categories of time, error distance, pattern of search, and amount of stereotypy are factors that describe the good problem solver when compared with the poor problem solver.

2. *Analysis of time scores.* The purpose of an analysis of time scores was to determine whether time is a significant factor in problem-solving proficiency. Stated in another way, does the good problem solver require significantly less time than the poor problem solver, when gear-train malfunctions are to be located?

Table 2 shows the means and standard deviations of time in minutes required by each of three problem-solving groups to make a decision with respect to the location of each of a series of six problems. Arriving at a decision was not nec-

TABLE 3

ANALYSES OF VARIANCE OF TIME IN SECONDS REQUIRED TO DECIDE THE LOCATION OF SOLUTIONS TO PRE-, INTER-, AND POSTTEST PROBLEMS

Test Phase	Source of Variance	All Problems			Problems Solved			Problems Failed		
		df	Mean Square	*F*	*df*	Mean Square	*F*	*df*	Mean Square	*F*
Pretest	Between groups	2	140,639.50	1.44	2	513.03	0.20	2	1,000.46	0.26
	Within groups	46	97,715.52		42	2,562.06		46	3,901.38	
Intertest	Between groups	2	114,728.11	2.13	2	732.57	1.05	2	6,041.70	1.78
	Within groups	46	53,904.21		42	698.14		46	3,397.04	
Posttest	Between groups	2	25,613.26	0.87	2	930.00	3.18	2	115.64	0.04
	Within groups	46	29,393.75		44	292.80		45	2,710.05	
Total test	Between groups	2	2,484.00	2.22	2	929.42	2.53	2	1,056.27	0.52
	Within groups	46	1,116.78		46	367.03		46	2,020.73	

essarily indicative that the decision was a correct one. Data describing performance on the various test phases are shown in Table 2 in the form of means and standard deviations for (*a*) all problems, whether passed or failed, (*b*) only those problems correctly solved, and (*c*) only those problems failed.

An inspection of the mean times in Table 2, both under problems solved and problems failed, shows a relationship between response latency (time in minutes to arrive at a decision) and problem-solving success. This means simply that all *S*s tended to require less time to arrive at correct solutions than to arrive at incorrect solutions.

In order to determine whether these two differences among the groups were statistically different from zero a set of analyses of variance was computed.

More specifically, 12 analyses of variance were computed. This number was arrived at by testing for differences among (*a*) all problems, (*b*) problems solved correctly, and (*c*) problems failed for each of the test phases (pre-, inter-, post-, and total test). The results of these analyses are summarized in Table 3.

Each of the 12 *F* ratios failed to reach significance at the 5 per cent level of confidence. The consistent lack of significant time difference among problem-solving groups is highly suggestive that time is a dubious factor in the description of the comparative properties of good and poor problem solvers.

3. *Analysis of error-distance dimension.* The purpose of analyzing error-distance scores was to study the relationship between this particular response dimension and problem-solving proficiency. More specifically, it was desired to discover whether more proficient problem solvers, although incorrect in their decision as to the exact location of a malfunction, were consistently less in error distance than the less proficient problem solvers.

Table 4 shows the means and standard deviations of error distances for the three problem-solving groups with respect to three series of six problems each. These error distances are expressed in terms of distance of response given by *S* from the functional unit in which the malfunction was actually located.

Thus, error distance was a direct function of the discrepancy (in terms of FU's) between indication of the location of a malfunction and the correct location of that malfunction. When the location indicated by *S* and the correct location were identical, the error-distance score was, of course, zero.

The mean error distances for the three groups show a steady decrease for the pre-, inter-, and posttest series of prob-

TABLE 4

COMPARISON OF ERROR DISTANCE RESPONSES MADE BY *Ss* CLASSIFIED ACCORDING TO PROBLEM-SOLVING PROFICIENCY

Means and Standard Deviations

Test Phase	Group I		Group II		Group III	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Pretest	13.6	7.22	31.0	13.02	31.1	10.78
Intertest	12.4	8.35	24.0	10.67	21.6	9.99
Posttest	4.9	3.79	9.3	8.90	19.7	10.08
Total test	30.9	12.60	64.3	22.08	72.4	23.68

Analysis of Variance Results

Test Phase	Source of Variation	*df*	Mean Square	*F*
Pretest	Between groups	2	1,376.48	11.06**
	Within groups	46	124.43	
Intertest	Between groups	2	445.06	4.32*
	Within groups	46	103.04	
Posttest	Between groups	2	994.55	12.72**
	Within groups	46	78.21	
Total test	Between groups	2	6,910.31	14.75**
	Within groups	46	468.56	

* $p < .05$.
** $p < .01$.

lems respectively. Group II problem solvers (initially poor, but improved) demonstrated the sharpest reduction in error distance between the pre- and posttest measures. The mean error distances for the total series of 18 problems show Group III (poor problem solvers) to possess the highest average total error distances and Group I (good problem solvers) to possess the least average error distance for the total series. It appears that an inverse relationship exists between problem-solving proficiency and error distance. In other words, magnitude of error appears directly related to problem-solving proficiency. In order to test whether these differences in error distance were statistically different from zero, analyses of variance were computed for each of the test phases (pre-, inter-, post-, and total test). All *F* ratios, as shown in Table 4, were significant at the 5 per cent level of confidence or beyond.

A further illustration of the obtained relationship between error distance and correct solution is shown in Table 5, which summarizes the results of the product-moment correlation coefficient computed between error distance and pass-fail performance. The three negative correlation coefficients of −.61, −.68, and −.79 represent the relationships for the pre-, inter-, and posttest phases respectively.

TABLE 5

PRODUCT-MOMENT CORRELATION BETWEEN ERROR DISTANCE AND PASS-FAIL PERFORMANCE

Test Phase	Product-Moment
Pretest	−.61
Intertest	−.68
Posttest	−.79

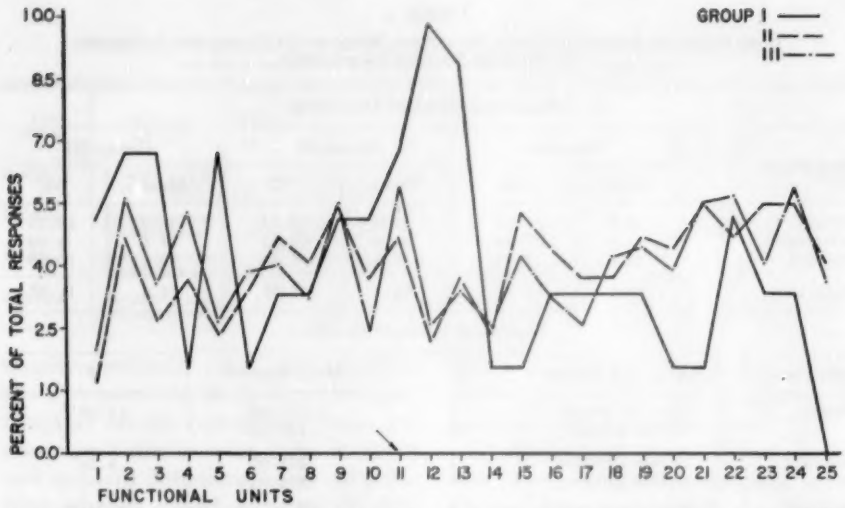

FIG. 2. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 11.

Since "error distance" is used as a simple linear concept in the present study, one interpretation of the foregoing coefficients of correlation is that the error distance for the good problem solvers was less, while for the poor problem solvers it was greater. The magnitude of these negative coefficients tends to increase from the pretest through the posttest.

4. *Analysis of "search pattern" profiles.* It has often been assumed that a good problem solver adopts a different way of attempting a solution than does the poor problem solver. In effect, the search patterns of good and poor problem solvers have been thought to differ.

The purpose, then, of an analysis of search patterns for the three problem-solving groups was to determine the relationship between this response dimension and problem-solving proficiency on the gear-train apparatus.

The gear train, it will be remembered, was divided into 25 functional units for the purpose of studying the combined search pattern of the three problem-solving groups. The meaning of search pattern in the present context is the frequency with which each group moved to each of the 25 FU's when the malfunction was located in only one of the 25 FU's. The present analysis, however, was not concerned with the *order* in which these responses occurred. The primary objective was simply to determine the relationships, by use of correlational methods, among the three response profiles for each of the 18 malfunctions, respectively.

An examination of Figure 2 provides a concrete illustration of what is meant. Figure 2 shows the comparative frequency, in terms of percentages of total responses, with which the three groups of problem solvers moved to each of the 25 FU's in the gear train, when malfunction number 1 was inserted in unit 11. This location is indicated by a small arrow along the abscissa. As is evidenced

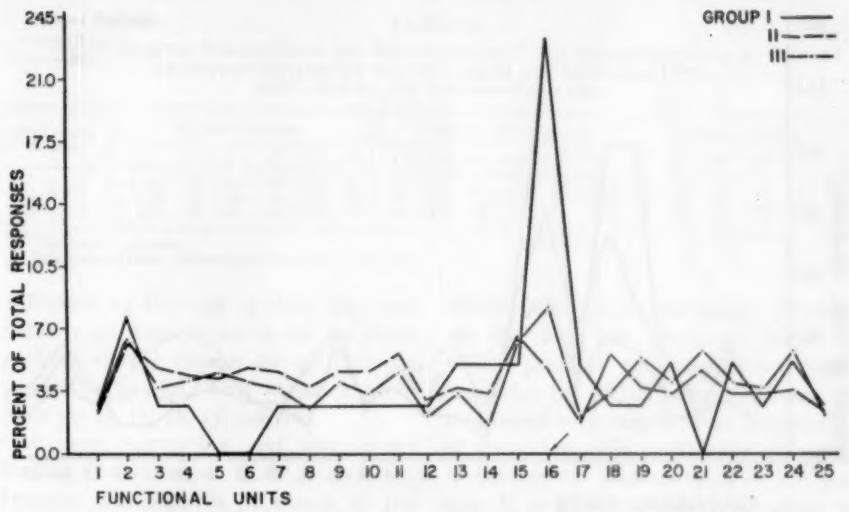

FIG. 3. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 16.

by the profiles, Group I (the good problem solvers) tended to group a greater percentage of their responses in the vicinity of functional unit 11 than did the remaining two groups. Product-moment coefficients of correlation were computed between the three groups, taken two at a time. Specifically, the correlation coefficient was computed between the percentage of their total responses that each group gave to each of the 25 FU's. The distribution of these percentages was in this way compared for the three groups, taken two at a time. Negative coefficients of correlation were obtained between Groups I and II, and I and III. The magnitude of these coefficients was −.33 (I and II) and −.09 (I and III) respectively. However, the correlation coefficient between Groups II and III was .76.

With respect to problem number 1 on the pretest, these coefficients indicate that no positive correlation exists between the pattern of search employed by the good problem solvers and the pattern displayed by the poorer problem solver. That Groups II and III, both classed as poor problem solvers at this stage, highly resembled each other in search pattern is evidenced by the correlation coefficient of .76.

Similarly, the correlation coefficients were computed between the percentage of their total responses that each group gave to each of the 25 FU's, for each of 18 malfunctions, the groups again being taken two at a time.

Fifty-four product-moment correlation coefficients were computed; the results are shown in Table 6.

Figures 3 through 7 show the comparative search patterns for the three problem-solving groups on the remaining pretest problems. The search patterns for pretest problem number 1 have already been shown in Figure 2.

Figures 8 through 13 show the search patterns for the three problem-solving groups during the intertest phase, which included problems 7, 8, 9, 10, 11, and 12.

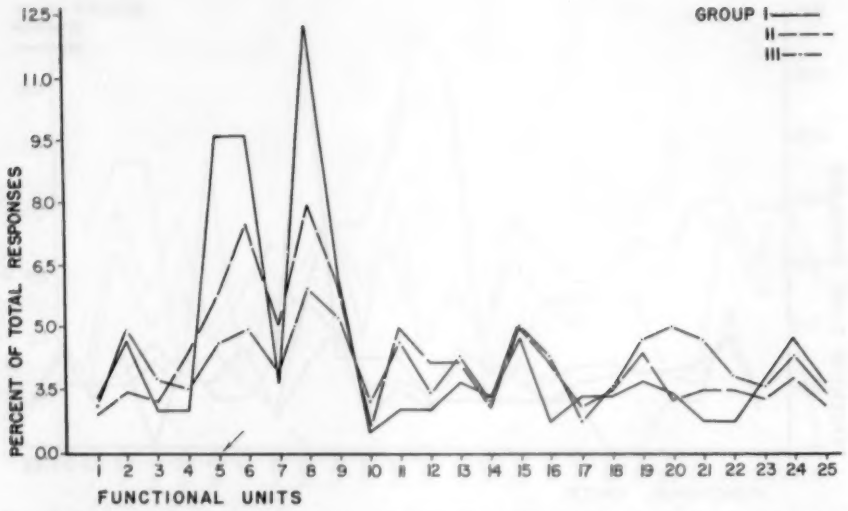

FIG. 4. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 5.

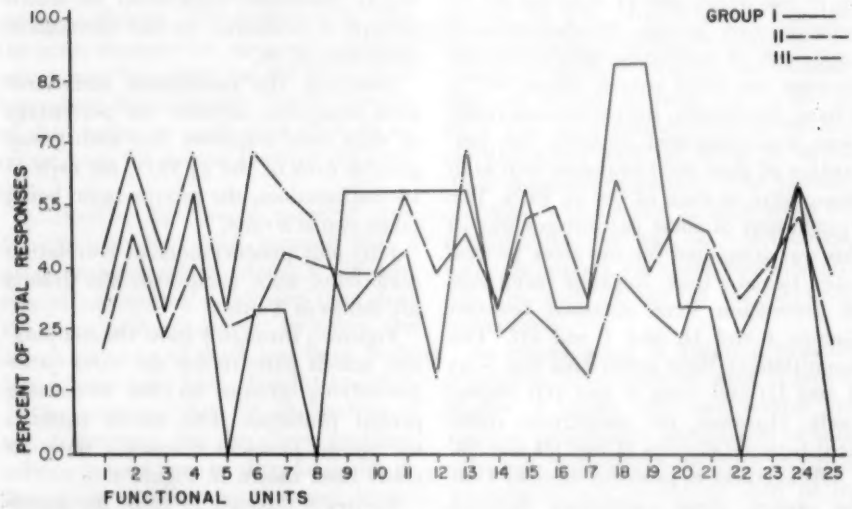

FIG. 5. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when actually no "malfunction" was inserted.

TABLE 6

PRODUCT-MOMENT CORRELATIONS FOR PERCENTAGE OF TOTAL RESPONSES GIVEN BY THREE GROUPS OF PROBLEM SOLVERS TO EACH OF 25 FUNCTIONAL UNITS (PRE-, INTER-, AND POSTTEST PROBLEMS)

Proficiency Groups	Pretest Problems						Intertest Problems						Posttest Problems					
	1	2	3	4*	5†	6	7	8	9	10*	11†	12	13	14	15*	16†	17	18
I vs. II	−.33	.36	.79	.56	.31	.83	.39	.50	.61	.63	.36	.76	.89	.77	.04	.53	.93	.97
I vs. III	−.09	.53	.56	.14	.29	.52	.60	.53	.57	.63	.07	.01	.82	.55	.15	.53	.81	.85
II vs. III	.76	.62	.72	.12	.72	.30	.67	.66	.59	.40	.63	.72	.81	.50	.47	.57	.80	.77

* "No defect" problem.
† Extremely difficult malfunction in the differential gears.

Figures 14 through 19 show the comparative patterns of search for the three problem-solving groups for each of the posttest problems. This includes problems 13, 14, 15, 16, 17, and 18.

A most interesting and unexpected finding is the trend of the correlation between the patterns of search of the three problem-solving groups to *increase*. An examination of Figures 2 through 19 indicates a general tendency for the search patterns between the groups to become highly similar. The search patterns for these same groups were extremely different at the outset, as shown on the first few problems. Table 6, which presents the correlation coefficients for the search patterns, points up this trend very clearly. The magnitudes of the correlation coefficients are much higher on the posttest than on the pretest. It is concluded that, in terms of behavioral movements, the poor problem solvers tend to resemble the more proficient problem solvers after a number of trials. In effect, the poor problem solvers have, as they practiced, adopted a search pattern that is highly

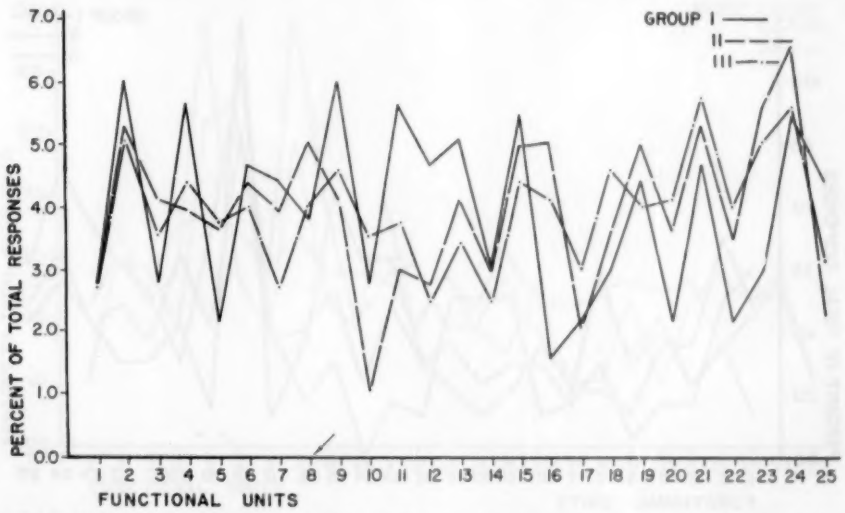

FIG. 6. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 8.

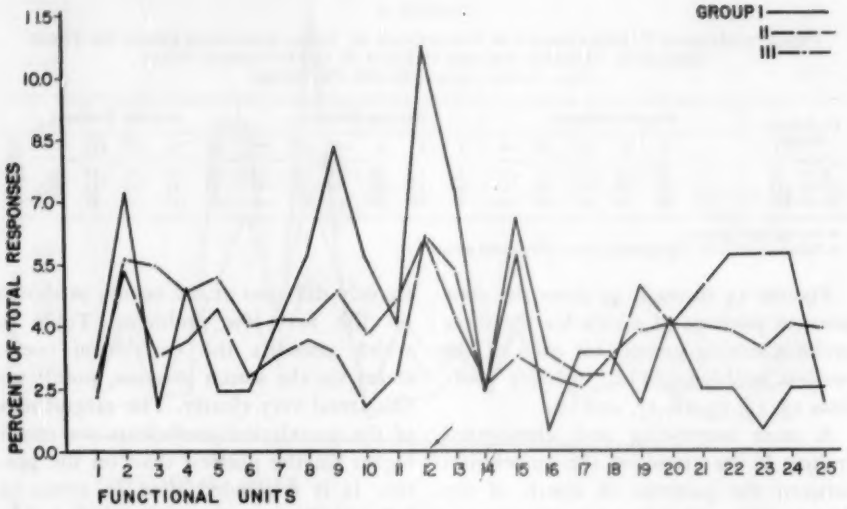

FIG. 7. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 12.

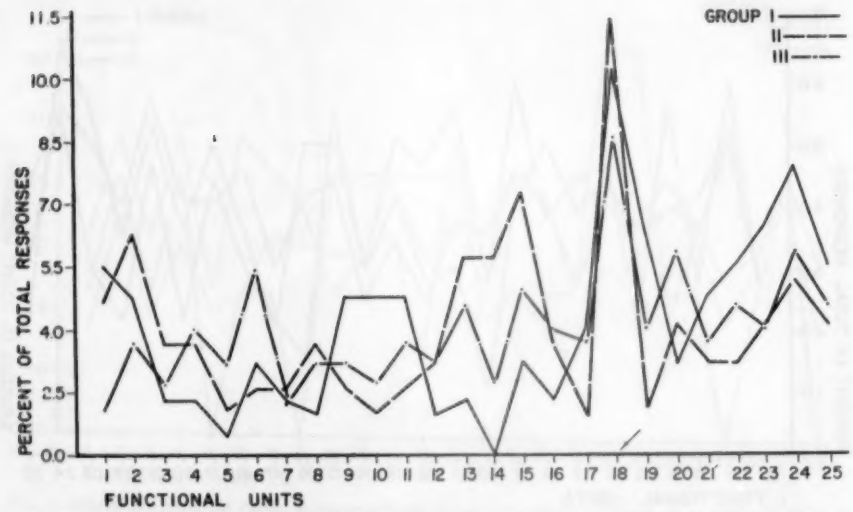

FIG. 8. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 18.

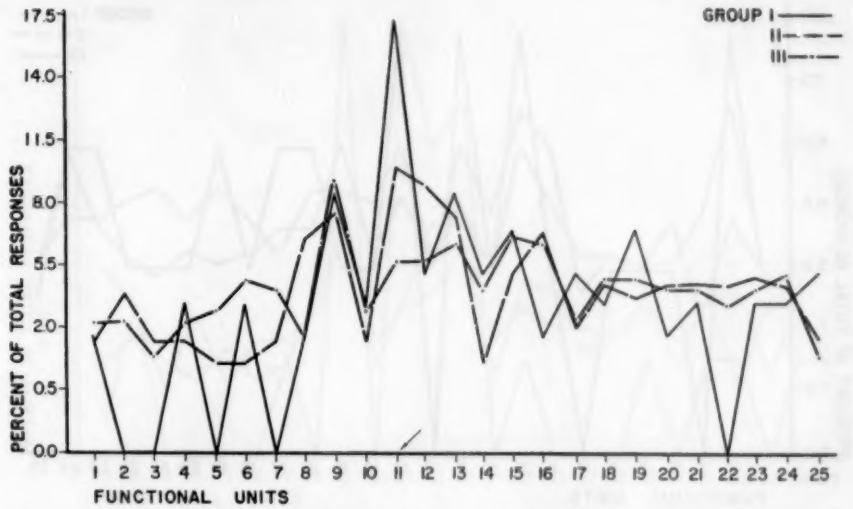

FIG. 9. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 11.

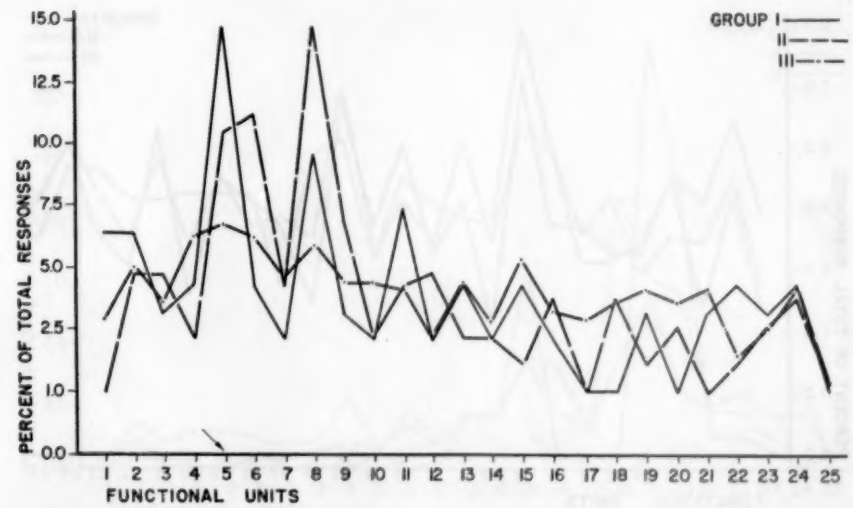

FIG. 10. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 5.

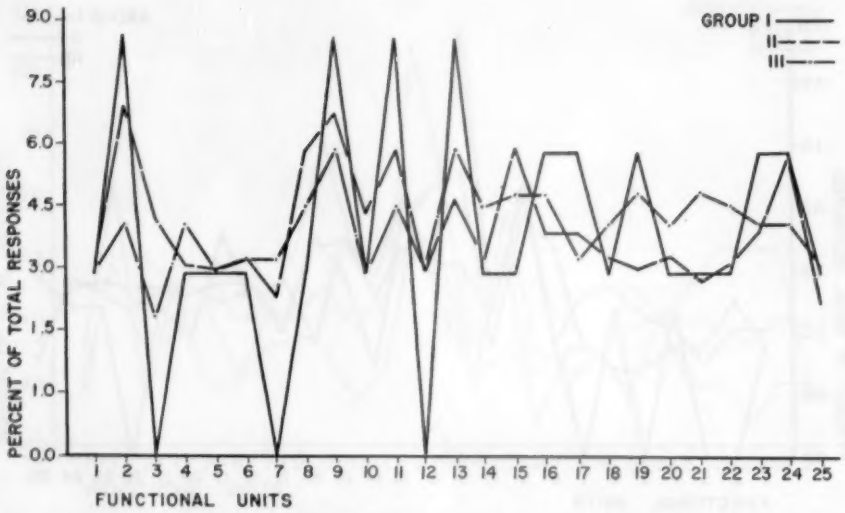

FIG. 11. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when actually no "malfunction" was inserted.

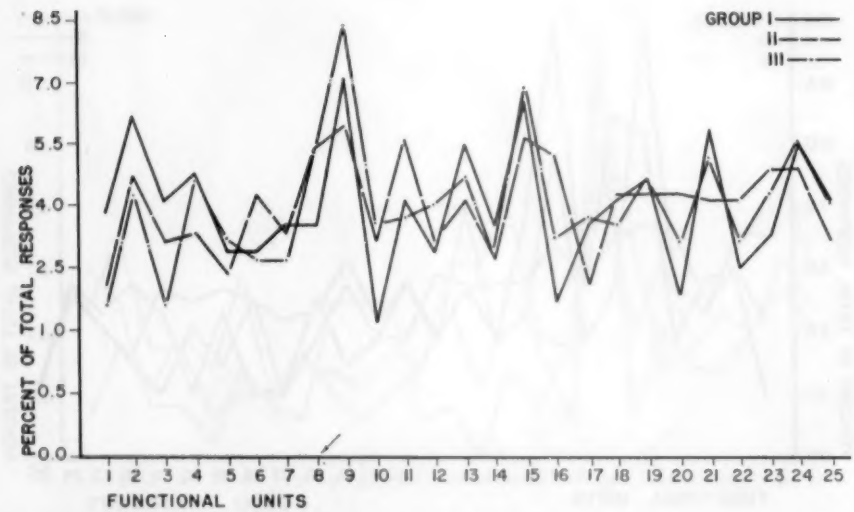

FIG. 12. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 8.

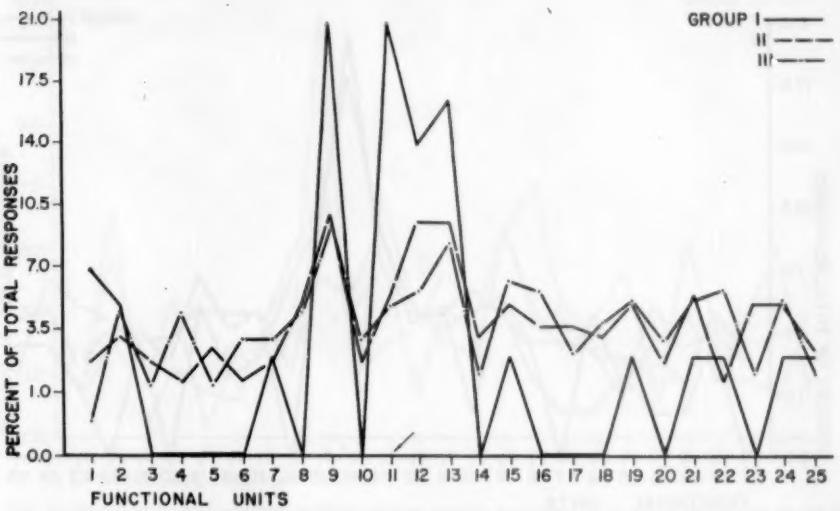

FIG. 13. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 11.

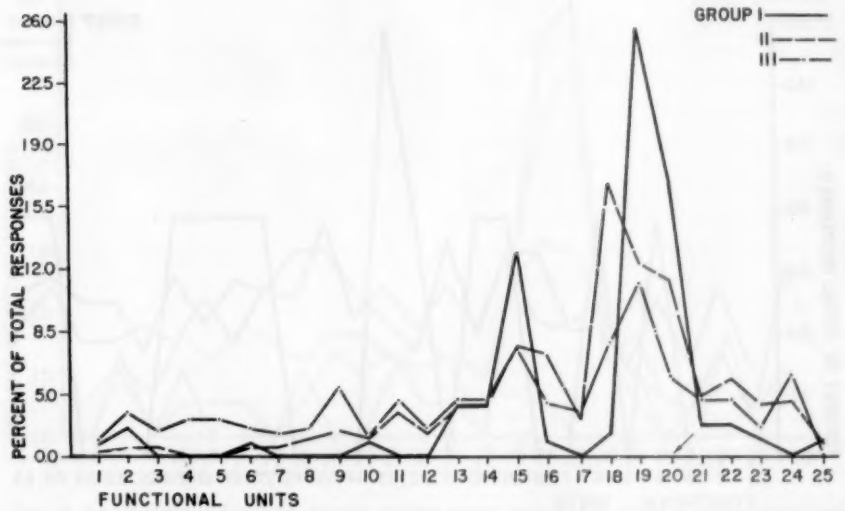

FIG. 14. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 20.

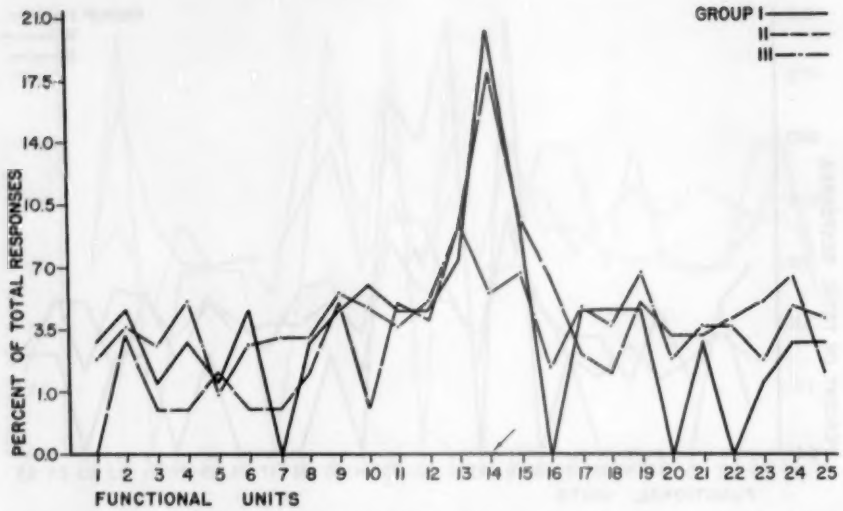

FIG. 15. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 14.

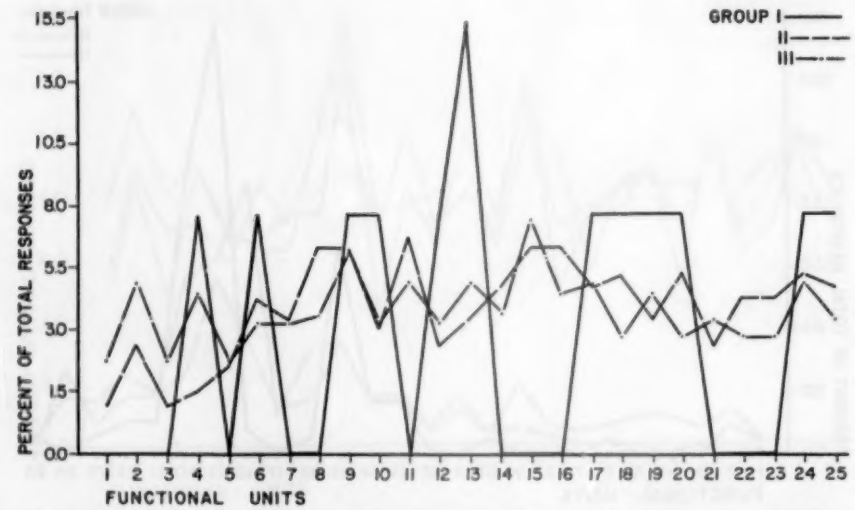

FIG. 16. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when actually no "malfunction" was inserted.

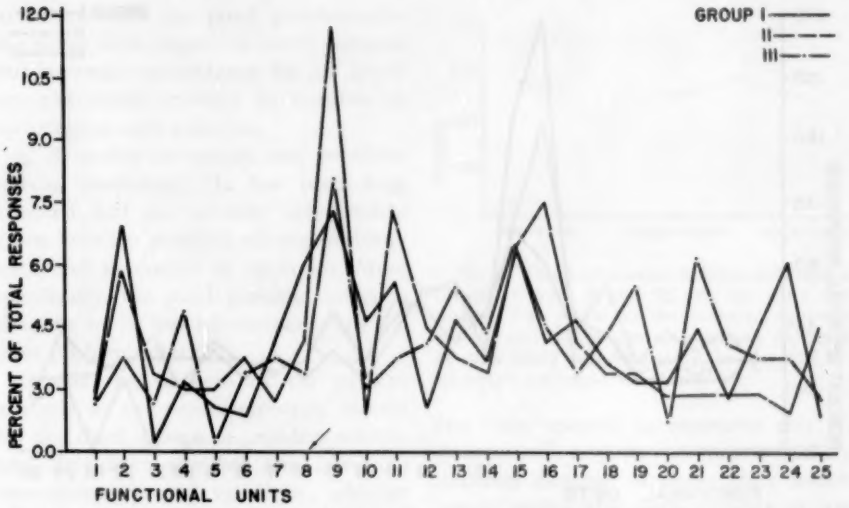

FIG. 17. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 8.

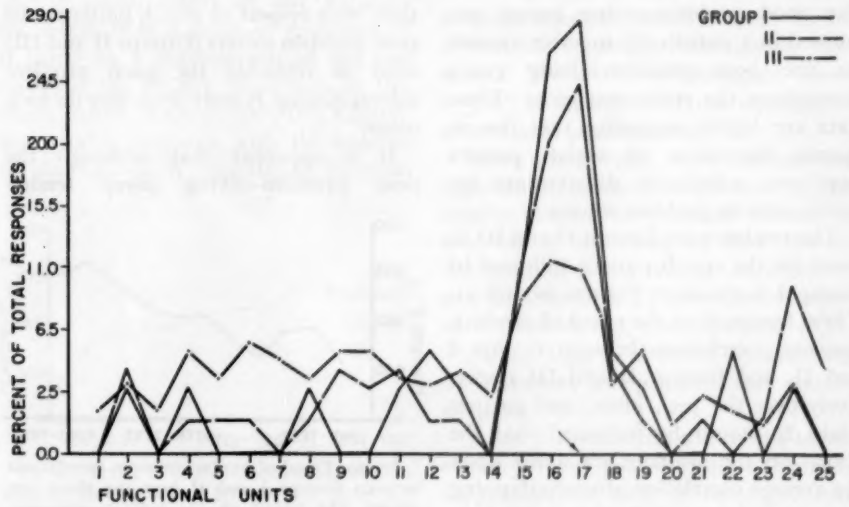

FIG. 18. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 17.

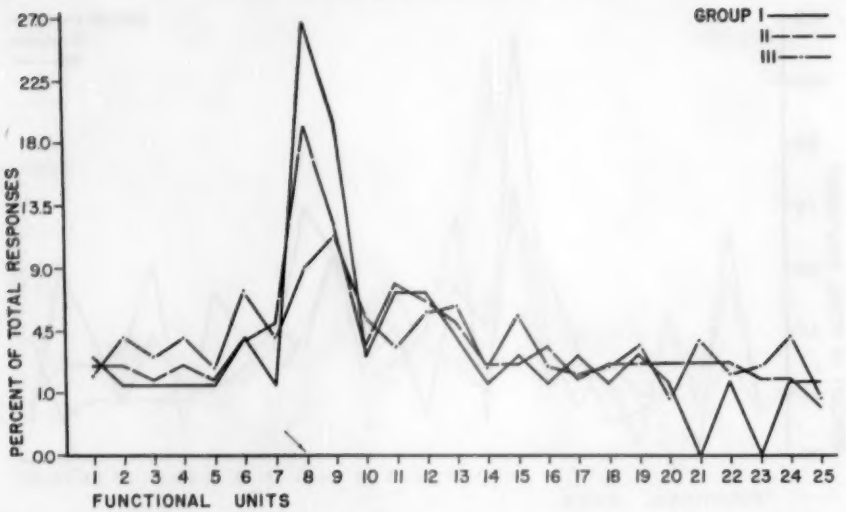

FIG. 19. Search patterns for three problem-solving groups expressed in terms of percentage of total responses made by each group to each of 25 gear-train units, when the "malfunction" is located in unit number 8.

similar to the search pattern used by the good problem solvers.

However, in terms of correct solutions the good problem-solving group performed in a statistically superior manner to the poor problem-solving group throughout the entire test series. These data are highly suggestive that the response dimension of search pattern does not sufficiently differentiate between various problem solvers.

The tendency for Groups II and III to resemble the search pattern followed by Group I is shown in Figures 20 and 21. These figures show the trend of product-moment correlations between Groups I and II, and Groups II and III respectively over the pre-, inter-, and posttest trials. It should be indicated that the points on the ordinate are based upon the average correlation after overlapping sequences of three trials each were taken. This was done to reduce the trial-by-trial variation. Figure 22 shows correlational trend data between Groups II and III.

It is apparent from these trend graphs that, with respect to search patterns, the poor problem solvers (Groups II and III) tend to resemble the good problem solvers (Group I) more than they do each other.

It is apparent that, although the poor problem-solving group tended

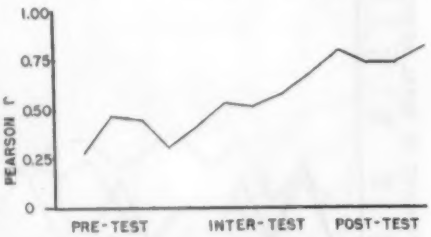

FIG. 20. Trend of product-moment correlations between Groups I and II over the three test phases. The points on the ordinate represent *average* correlations after overlapping sequences of three trials each were taken, so that trial-by-trial variation might be reduced.

to behave like the good problem-solving group with respect to search pattern, this increased resemblance did not imply any noticeable increase in number of correct gear-train solutions.

5. *Response stereotypy and problem-solving proficiency.* It has been long assumed that an "inverse" relationship exists between problem-solving performance and stereotypy of approach. More specifically, the good problem solver is thought to be less stereotyped than the poor problem solver.

It was the purpose of the present analysis to compare stereotypy indices for the three groups of problem solvers, and determine whether the foregoing assumptions proved valid, i.e., whether good problem solvers are indeed *less* stereotyped than poor problem solvers.

The index of behavioral stereotypy as developed by Miller and Frick (2, 3) was used in the analysis.

This index of stereotypy is a measure of the predictability of any sequence (of length s) in a chain of responses. It is based on a characteristic of any one response (or of any sequence) in this chain, called its *uncertainty* (U). Clearly, this quality increases with the number of alternatives. A logarithmic function

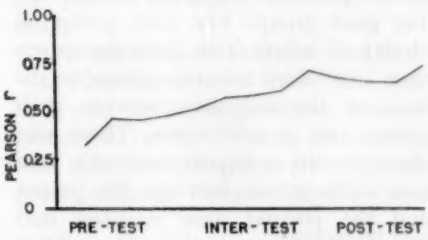

FIG. 21. Trend of product-moment correlations between Groups I and III over the three test phases. The points on the ordinate represent *average* correlations after overlapping sequences of three trials each were taken, so that trial-by-trial variation might be reduced.

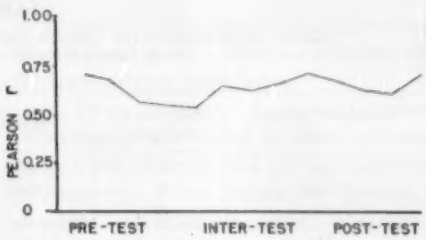

FIG. 22. Trend of product-moment correlations between Groups II and III over the three test phases. The points on the ordinate represent *average* correlations after overlapping sequences of three trials each were taken, so that trial-by-trial variation might be reduced.

has been selected to represent this U, because it is necessary to have a directly additive measure of uncertainty which would assign the same amount of uncertainty to all successive choices from a alternatives, assuming all choices to be equally probable. $U(a)$ is derived to be $K\log_n(a)$, where a is the number of alternatives. Any case of unequal probabilities is then proved to be reducible to a case of equal probabilities by the insertion of a proper multiplying constant. Thus, $U(a) = cK\log_n(a)$, where $0 \leq c \leq 1$, regardless of the equality or otherwise of probabilities.

Miller and Frick have chosen the logarithmic function to the base 2 principally because they choose to talk of uncertainty in terms of *bits* (binary digits). This is not crucial, since the index of *stereotypy*, C_s, is defined as equal to 1—(relative U_s), where (relative U_s) = U_s/U_{max}. U_{max} is obtained by assuming all responses independent and equally frequent.

Table 7 summarizes the results of the stereotypy analysis, showing the stereotypy index for each of three problem-solving groups over the pre-, inter-, and posttest phases. Stereotypy indices are shown also for total test performance.

TABLE 7

GEAR-TRAIN STEREOTYPY INDICES FOR SEQUENCES OF 1, 2, AND 3 RESPONSES FOR THREE GROUPS OF PROBLEM SOLVERS

Problem-Solving Group	Sequence Length	Pretest	Intertest	Posttest	Total Test
I (Good)	1	.0077	.0068	.0058	.0025
	2	.2942	.2424	.2220	.2451
	3	*	*	*	.2668
II (Poor-Improved)	1	.0089	.0021	.0148	.0038
	2	.3535	.2529	.2683	.2807
	3	*	*	*	.3025
III (Poor-Unimproved)	1	.0130	.0055	.0037	.0068
	2	.3601	.3578	.3232	.3453
	3	*	*	*	.2979

* Insufficient data.

The asterisks in Table 7 indicate that an adequate number of responses was not available to compute sterotypy indices for sequences of three responses. Therefore, for sequences of three responses it was possible to compute only the total test stereotypy index for each of the three groups. To compute sequences of three responses for the pre-, inter-, and posttest separately would have required in *excess* of 10,000 responses for each phase.

The trend in stereotypy indices is for the good problem solvers to have a lower index than the poor problem solvers. These data lend quantitative weight to the assumption that poor problem solvers are more repetitive in their movements when compared with good problem solvers. These data are highly suggestive of the conclusion that stereotypy of behavior is a response dimension that appears promising in distinguishing between levels of problem-solving proficiency.

SUMMARY AND CONCLUSION

Progress in establishing relationships between environmental and response variables in problem solving demands first that attention be directed toward development of response measures *other* than time and number of correct solutions. A step in this direction was taken by comparing good and poor problem solvers on three additional response dimensions.

The investigation was divided into a pretest, intertest, and posttest phase. Forty-nine adult *Ss* were individually given six problems to solve during each of these foregoing phases. It was first necessary to demonstrate that *Ss* classified as good problem solvers were more proficient than those classified as poor. Solution of three or more of the six pretest problems was a requirement for the good group. The poor group included all others. This latter group was then subdivided into two groups on the basis of the difference between their pretest and posttest scores. Those who demonstrated an improvement of at least two malfunctions between the pretest and the posttest were classified into Group II, "poor improved" problem solvers. The remaining subjects were placed in Group III and referred to as "poor unimproved" problem solvers. Differences between groups were estab-

lished beyond the 1 per cent level of confidence.

A gear train, consisting of two series and four parallel channels into which problems were easily inserted, was the apparatus.

Failure of the panel to illuminate after the control handles were operated indicated to *S* that the apparatus was defective; the problem was to find the defect. Training factors were common to all *S*s. Standard operating procedure instructions were given before the pretest. Immediately following the pretest, a tape-recorded basic knowledge lecture concerning the nomenclature of the gear train was given to all *S*s. This was followed immediately by the intertest. After a 24-hour interval a problem-solving lecture specific to the gear train was administered. This problem-solving lecture was given just prior to the posttest. All 66 moving parts of the apparatus were numbered, and *S*s were required to call out the numbers of all parts as they went from part to part in the process of attempting a solution.

In addition to time measures, the groups were compared with respect to (*a*) error distance, (*b*) stereotypy of response on each trial over 18 trials, and (*c*) "pattern of search" for each of 18 trials.

The following results were obtained:

1. *Time.* Analyses of time data were computed for (*a*) all problems, whether passed or failed, (*b*) only those problems correctly solved, and (*c*) only those problems failed. None of the 12 *F* ratios so obtained were significant, suggesting that time may be a dubious criterion of problem-solving proficiency.

2. *Error distance.* A strong inverse relationship was found to exist between magnitude of error and problem-solving proficiency. The following product-moment correlations were obtained: pretest, −.61; intertest, −.68; posttest, −.79.

3. *Pattern of search.* Correlation coefficients between group patterns *increase* as a function of number of trials. Poor problem solvers "searched" increasingly like good ones, but remained inferior in performance.

4. *Response stereotypy.* The Miller-Frick index indicated that the poor problem solvers are consistently more stereotyped in their problem-solving behavior than the good problem solvers. Whether these differences were "significant" is unknown since the authors knew of no existing test of significance to apply.

REFERENCES

1. Fattu, N. A., & Mech, E. V. The effect of set on performance in a "trouble shooting" situation. *J. appl. Psychol.*, 1953, **37**, 214-217.
2. Frick, F. C., & Miller, G. A. A statistical description of operant conditioning. *Amer. J. Psychol.*, 1951, **64**, 20-36.
3. Miller, G. A., & Frick, F. C. Statistical behavioristics and sequences of responses. *Psychol. Rev.*, 1949, **56**, 311-324.
4. Tresselt, M. E., & Leeds, D. S. The effect of concretizing the mental set experiment. *J. gen. Psychol.*, 1953, **48**, 51-55.
5. Underwood, B. An orientation for research on thinking. *Psychol. Rev.*, 1952, **59**, 209-220.

(Accepted for publication March 1, 1954)

GEORGE BANTA PUBLISHING COMPANY, MENASHA, WISCONSIN